

Accepted Manuscript

A machine-learning based solution for chatter prediction in heavy-duty milling machines

Ibone Oleaga, Carlos Pardo, Juan J. Zulaika, Andres Bustillo

PII: S0263-2241(18)30554-2
DOI: <https://doi.org/10.1016/j.measurement.2018.06.028>
Reference: MEASUR 5645

To appear in: *Measurement*

Received Date: 15 May 2018
Revised Date: 14 June 2018
Accepted Date: 15 June 2018



Please cite this article as: I. Oleaga, C. Pardo, J.J. Zulaika, A. Bustillo, A machine-learning based solution for chatter prediction in heavy-duty milling machines, *Measurement* (2018), doi: <https://doi.org/10.1016/j.measurement.2018.06.028>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A machine-learning based solution for chatter prediction in heavy-duty milling machines

Ibone Oleaga¹, Carlos Pardo², Juan J. Zulaika¹, Andres Bustillo^{2,*}

¹ Industry and Transport Division, Tecnalia, San Sebastian, Spain

² Department of Civil Engineering, University of Burgos, Burgos, Spain

*Corresponding author: Department of Civil Engineering, University of Burgos, Avda Cantabria s/n, 09006, Burgos, Spain, telephone: +34 947259358, fax: +34 947258910, email: abustillo@ubu.es

Abstract

The main productivity constraints of milling operations are self-induced vibrations, especially regenerative chatter vibrations. Two key parameters are linked to these vibrations: the depth of cut achievable without vibrations and the chatter frequency. Both parameters are linked to the dynamics of machine component excitation and the milling operation parameters. Their identification in any cutting direction in milling machine operations requires complex analytical models and mechatronic simulations, usually only applied to identify the worst cutting conditions in operating machines. This work proposes the use of machine-learning techniques with no need to calculate the two above-mentioned parameters by means of a 3-step strategy. The strategy combines: 1) experimental frequency responses collected at the tool center point; 2) analytical calculations of both parameters; and, 3) different machine-learning techniques. The results of these calculations can then be used to predict chatter under different combinations of milling directions and machine positions. This strategy is validated with real experiments on a bridge milling machine performing concordance roughing operations on F-114 steel with a 125 mm diameter mill fitted with nine cutters at 45°, the results of which have confirmed the high variability of both parameters along the working volume. The following regression techniques are tested: artificial neural networks, regression trees and random forest. The results show that random forest ensembles provided the highest accuracy with a statistical advantage over the other machine learning models; they achieved a final accuracy of 0.95 mm for the critical depth and 7.3 Hz for the chatter frequency (RMSE) in the whole working volume and in all feed directions, applying a 10x10 cross validation scheme. These RMSE values are acceptable from the industrial point of view, taking into account that the critical depth of this range varies between 0.68 mm and 19.20 mm and the chatter frequency between 1.14 Hz and 65.25 Hz. Besides, random forest ensembles are more easily optimized than artificial neural networks (1 parameter configuration versus 210 MLPs). Additionally, tools that incorporate regression trees are interesting and highly accurate, providing immediately accessible and useful information in visual formats on critical machine performance for the design engineer.

Keywords: Random Forest, regression trees, milling, vibrations, chatter, polar diagrams

Nomenclature

X, Y, Z_{MT}	X, Y, Z-axis of the Machine-tool
X, Y, Z_t	X, Y, Z-axis of the tool tip
Z	Number of flutes of the tool
κ	Tool cutting-edge lead angle
\bar{I}_m	Unit vector that defines the direction of a m^{th} machine mode
$\{\phi_i\}$	i -Modal vector per unit of mass
\bar{I}_m^{xy}	Projection of the unit vector over the $X_{MT}Y_{MT}$ plane
θ_{xy}	Angle between \bar{I}_m^{xy} and $+X_t$
α_x	Angle between \bar{I}_m^{xy} and $+X_{MT}$
α_f	Angle between $+X_t$ and $+X_{MT}$
θ_z	Angle between $\{\phi_i\}$ and the $X_{MT}Y_{MT}$ plane
$\{\delta\}$	Modal displacements
h_d	Chip thickness
$\{g\}$	Dynamic forces in modal coordinates
K_t	Tangential cutting force coefficient for the tool and workpiece material
a_p	Axial depth of cut
a_{pcrit}	Critical depth of cut
$\{\Delta\delta\}$	Dynamic displacements of the tool in modal coordinates.
ω_c	Chatter frequency
τ	Tooth passing period
ξ_i	Damping coefficient for the chatter frequency ω_i
k_{efi}	Effective stiffness of the i -mode at the tool center point
ϑ	Machining feed direction

1. Introduction

The productivity and the surface quality of milling operations are limited by many process parameters, from cutting conditions and surface topography [1] to tool orientation [2] and cutting forces [3]. But, in the great majority of cases they are constrained by self-induced vibrations [4,5], especially regenerative chatter vibrations, caused by the regeneration of the chip thickness in systems where the cutting edge cuts a previously-machined surface and generates forces that excite machine component modes of lower dynamic stiffness. Those modes can be due either to structural components of the machine [6,7] (usually of low frequency, below 200 Hz) or to the spindle-toolholder-tool system [8] (at higher frequencies, above 200 Hz), or to the workpiece when it has flexible thin walls or floors (above 500 Hz) [9]. The excitation frequencies linked to the

milling operations are within a range that is below 100 Hz in the specific case of machining hard materials such as tempered and hardened steels, so, the stability of the milling processes in the case of heavy-duty milling machines is highly dependent on the modal parameters of the machine structure. For these cases, the critical modes of the process-machine interaction are therefore linked to the structural modes of the milling machine.

Concerning regenerative chatter vibrations, the key parameters in the case of regenerative chatter vibrations are the vibration frequency and the cutting depth at which the regenerative process occurs; the first parameter can be changed by the process engineer before the machining program started while the second is interesting for the machine-tool designer to improve the machine design for future machines. Determining these two key parameters of chatter frequency and of available depth of cut requires an analysis of the following parameters: the workpiece material (cutting force coefficients), the tool type (geometry of the tool, number of teeth, cutting edge lead angle), the machining parameters (radial and axial immersion of the tool, rotating speed, feed rate etc.) and the modal parameters of the machine-workpiece system [10]. These parameters are all related to the machining operation except for the modal parameters of specific machining positions that are part of the design of the machine and cannot be altered. Therefore, if hard materials are to be machined, besides conceiving stiff and robust machines at the design stage, complex stability models that incorporate experimental data should also be considered at that stage. Thus, the dynamics of the machine structure at the Tool Center Point (TCP), the dynamics of milling forces and the dynamics of chip thickness will be taken into consideration as regards the operational stage of the machine.

Machine-learning techniques can be used to predict chatter-frequency and depth-of-cut, because these models are able to extract information from noisy environments [11]. A major problem arises here: the small size of the experimental datasets recorded under factory conditions. This problem arises from real restrictions in an industrial setting: first, process engineers in a factory are unwilling to test many different milling conditions, due to the extra-cost that these experiments imply; and, secondly, the only way to describe the physical process properly is to position new sensors, mainly accelerometers [12], on the machine-tools to supply more information on the process performance. This requirement clashes with the strict objective of cost-reduction in modern industry. The immediate conclusion of these limitations is that real datasets for chatter prediction will be poor. Besides, the dynamic behavior, and therefore the chatter frequency and the depth of cut of milling machines and the positioning of the milling head [13] change in each new machine design, complicating the collection of up-to-date datasets. This situation means that, up until the present, there has been no attempt to model chatter frequency in heavy duty machines using machine-learning techniques, although two different approaches are presented in the bibliography for tool-chatter analysis that consider machine-learning techniques. The first approach uses machine-learning techniques to classify the current machining status; mainly “stable” and “unstable” cutting conditions [12, 14-18]. The second approach, very close to the aim of this research, and presented only two years ago in the literature [19,20] combines mathematical models with machine-learning techniques to predict the maximum stable cutting depth in real time.

Related to the first approach, different machine-learning algorithms have been tested to classify real-time machining status: Lamraoui et al [14] used Support Vector Machines (SVMs) at first only using motor current signals and then radial basis functions and Multi-Layer Perceptrons (MLPs) based on vibrational signals [12], to classify both stable and unstable cutting conditions. Kuljanic et al [15] used wavelet decomposition to extract statistical parameters from vibration signals collected by accelerometers on the machine head, feeding them into an Artificial Neural Network

(ANN). In this case, the wavelet decomposition method was preferred to Fourier transform. Yao et al [16] used the same strategy based on wavelet decomposition for vibrational signals but using SVMs for the state classification considering in this case three states: chatter, no-chatter and a transitional state; and fuzzy logic was used as the diagnosis algorithm by Bediaga et al [17] to classify stable and unstable high-speed milling. Finally, Sofuoglu et Orak [18] predicted a stable cutting depth in turning operations without chatter vibrations comparing different machine-learning algorithms: regression analysis, ANNs, and decision trees. They concluded that ANNs produced the best results when the algorithms were fed with information on different workpieces and insert materials: geometries, cutting and modal parameters obtained from real experiments on different materials (aluminum, inconel and steels).

Related to the second approach, involving a combination of mathematical models with machine-learning techniques to predict the maximum stable cutting depth in real time, Friedrich et al [19] compared the performance of SVMs and ANNs (radial basis functions) to predict chatter in a milling process feeding both models with the scenarios obtained from a simulation model that included: chip load, cutting force and machine-tool models with the final mathematical calculation of the stability lobe diagrams. Hu et al [20] presented a second attempt along these lines by using ANNs to predict the limiting axial cutting depth; a mathematical model of the stability lobes generated by the training data, while the ANN received 9 inputs to calculate the lobes: the tangential and normal cutting force coefficients, the modal damping and stiffness coefficients, the modal mass, damping and stiffness coefficients in orthogonal directions, and the radial immersion ratio.

Although there are few references related to machine-learning modeling of structural chatter, wear and surface prediction in machining processes are similar processes where these techniques have been tested with good results, as reported in two recent reviews [21-23]. These reviews also concluded that ANNs are often used as a baseline technique [21-23] and the new approaches try to improve the accuracy of these sorts of models [24-26]. Among these new approaches, the use of ensemble methods [27] in the manufacturing process have promising applications; ensembles use several classifiers at the same time, each classifier provides its prediction and all the predictions are joined. They have demonstrated their high prediction accuracy in different manufacturing processes: Yü [28] used ensembles to identify out-of-control signals in multivariate processes, Cho [29] and Bisaeid [30] used ensembles for end-milling condition monitoring and simultaneous detection of transient and gradual abnormalities in end milling, Bustillo et al. proposed the use of ensembles to predict surface roughness in face milling [24] operations and Liao et al. [31] and Bustillo and Rodriguez [25] used ensembles for grinding wheel and multitooth tool condition monitoring, respectively.

In this research, a strategy is proposed to side-step dataset limitations from the point of view of artificial intelligence. First, the FRFs are measured and both chatter frequency and critical depth are analytically calculated. Then the two critical parameters for structural chatter -depth of cut and chatter frequency- are calculated for each milling direction, so as to test different machine-learning techniques and to identify the most accurate and the cheapest in terms of their tuning requirements. Machine-learning experts are not regularly available in factories, so models that provide high accuracy without complex parametric tuning processes are preferable. Figure 1 shows the scheme of this strategy. Although similar schemes have been applied to other manufacturing processes such as surface roughness prediction [24], tool breakage detection [25], and surface wear prediction [26], it is validated here for the first time in this industrial task.

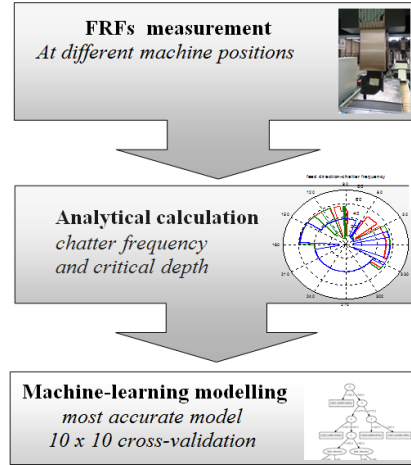


Figure 1. Scheme of the proposed methodology.

The remainder of this work is organized as follows: in the following section, the analytical determination of chatter frequency and critical depth will be presented; then, the different machine-learning techniques used to build the prediction models will be briefly described. Section 4 will discuss the experimental validation of the proposed methodology in a real heavy-duty machine. Finally, the conclusions and the future lines of research will be outlined in Section 5.

2. Analytical determination of chatter frequency and critical depth of cut

Mechanistic modeling of the milling forces and a model of the dynamic chip thickness as a function of vibrations in modal directions was used for the analytical calculation of chatter frequency and critical depth of cut [32]. A single-frequency approach was considered for the solution of the stability problem, which leads to the calculation of the chatter frequency ω_c [32]. Figure 2 shows the coordinate axes that were used and the directions of the machine modal vectors.

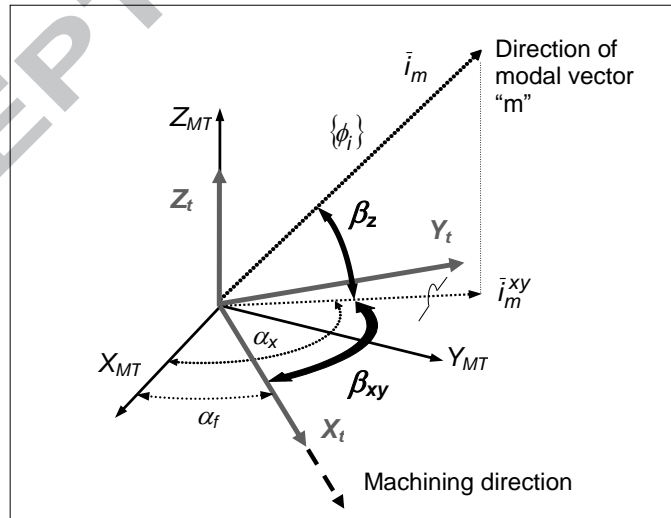


Figure 2. Directions of machine tool axes (MT indexes) and tool axes (t indexes) in a machine modal vector.

The machine axes, X_{MT} and Y_{MT} , define the working plane while the local axis $+X_t$ on the tool tip defines the feed direction of a tool of Z teeth and the cutting edge lead angle κ . \bar{l}_m is the unit vector that defines the direction of a m^{th} machine mode; $\{\phi_{ij}\}$ is the modal vector per unit of mass; \bar{l}_m^{xy} is the projection of the unit vector over the working plane, β_{xy} is the angle between that projection and the feed direction $+X_t$, and, α_x is the angle between that projection and the machine axis $+X_{MT}$; α_f is the angle between the feed direction $+X_t$ and the $+X_{MT}$ axis. Finally, β_z is the angle between the modal vector $\{\phi_{ij}\}$ and the $X_{MT}Y_{MT}$ working plane.

Using a mechanistic linear model, the dynamic forces are projected over the m^{th} mode direction \bar{l}_m , with these process forces expressed as modal coordinates. Modal displacements – referred to as $\{\delta\}$ – modulate the dynamic chip thickness h_d , in reaction to modal forces, in the terms expressed in Eq. (1).

$$\{g\} = K_t \cdot a_p \cdot [A(\phi)] \{\Delta\delta\} \cdot (1/2). \quad (1)$$

In Eq. (1), $\{g\}$ represents the dynamic forces in modal coordinates; K_t the tangential cutting force coefficient for the tool and workpiece material under consideration; a_p is the axial depth of cut; $[A(\phi)]$ is a matrix of dimensionless directional factors; and, $\{\Delta\delta\}$ represents the dynamic displacements of the tool in modal coordinates.

A mono-frequency solution is then applied, in order to solve the stability problem, so that if chatter appears at a frequency of ω_c , the dynamic displacements of an i^{th} mode in modal coordinates and in a tooth passing period, τ , depend on the dynamic cutting forces and on the modal parameters of that i^{th} mode: modal frequency ω_i , modal vector $\{\phi_{ij}\}$ and damping coefficient ξ_i . Also, the effective stiffness $k_{eff} = \omega_i^2 / \{\phi_{ij}\}^2$ of the mode at the Tool Centre Point (TCP) is obtained. Hence, Eq. (2) is obtained by taking the average Fourier term of the matrix of directional factors $[A(\phi)]$ and referring to $[\alpha]$ as the matrix of the average directional factors.

$$\{g\} = \frac{Z}{4\pi} \cdot K_t \cdot a_p \cdot (1 - e^{-i\omega_c \tau}) \cdot [\alpha] \cdot [H(\omega)] \cdot \{g\} \quad (2)$$

Equation (2) is an eigenvalue problem of the n^{th} order, where n is the number of modes under consideration; $[H(\omega)]$ is a diagonal matrix of the transfer functions, $h_i(\omega)$, at the tool tip in modal coordinates; Z is the number of flutes of the tool, and, τ is the tooth passing period.

The solution of the eigenvalue problem shown in Eq. (2) provides the relationship between the chatter frequency, ω_c , the depth of cut, a_p , and the spindle speed, shown in Eqs. (3) and (4). Charts known as stability lobes and chatter frequency diagrams may be constructed from the coefficients between those factors, as well as the maximum achievable depth of cut in the absence of chatter vibration $a_{p_{crit}}$, as in Eq. (5).

$$a_p = \frac{2\pi}{Z \cdot K_t \cdot \alpha \cdot \text{real}(h(\omega))} \quad (3)$$

$$\omega_c = \frac{N \cdot Z \cdot (\varepsilon + 2\pi \cdot n)}{60} \quad n=0,1,2... \quad (4)$$

$$a_{p_{crit}} \cong \frac{4\pi \cdot \sin \kappa \cdot k_{ef} \cdot \xi}{Z \cdot K_t \cdot |\sigma|} \quad (5)$$

In Eq. (3), α is the average directional factor, while in Eq. (5), σ is the directional factor, both of which are dimensionless factors that depend on the orientation between the mode and the feed direction. With regard to the critical depth of cut, for common tools and operations, the value is usually $|\sigma| < 1$. On the other hand, in Eq. (4), ε represents the diphas between the vibrations of two consecutive periods.

The representation of these equations considering the tool rotation speed, N , as an independent variable leads to the aforementioned stability lobes and chatter frequency diagrams. Figure 3 shows an example of these diagrams for the case of a face-milling operation on AISI 1045 steel with a tool diameter of 125 mm and nine teeth with a lead angle of 45° , machining on an X_{MT} direction with a Nicolas Correa Fox milling machine. The calculation of these two diagrams for any angular cutting direction are called polar diagrams for the critical depth and the chatter frequency.

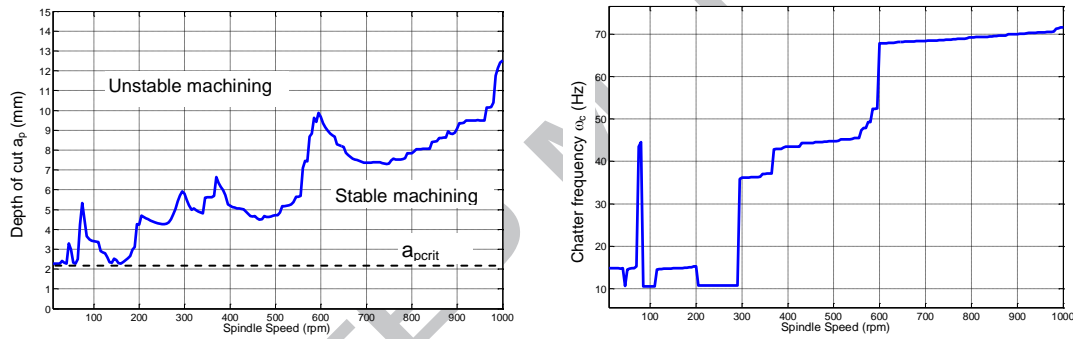


Figure 3. Examples of stability lobes (left) and chatter frequency diagrams (right) in a Nicolas Correa milling machine.

3. Machine-learning modeling

Three regression techniques of very different nature were tested using the WEKA data-mining tool [33]. First, the most common artificial neural networks, the Multilayer Perceptron (MLP) was tested. Multilayer Perceptron [34] is a kind of ANN, where the connections between neurons are weighted by a back-propagation algorithm. In this work, we used between 2 to 6 neurons in the hidden layer. The parameters momentum and learning rate were also tuned, as described in Section 4.2. MLPs were included in this research, because they are a standard in the modelling of many manufacturing processes by means of machine-learning techniques, such as surface roughness prediction [21,22] and tool wear prediction in cutting processes [23].

Regression trees were also included in this study, because they provide visual models of special interest for both machine-tool designers and process engineers. The M5P model tree [35] is a decision tree for regression. Decision trees consist of decision nodes connected by arcs. Each node has an input arc and two output arcs for numeric attributes (like the attributes considered in this

research), representing whether or not an attribute value of a given instance is greater than a given threshold. Final nodes have no output arcs and are called leaves. Each leaf includes a linear regression used to predict the output for all the instances that reach this leaf. The attributes and numeric thresholds in each node are picked when trying to subdivide the training instances into pure subsets (i.e.; subsets where almost all training instances are likely to fit the final linear model in the leaf). The M5P criterion for obtaining pure subsets is to maximize the expected error reduction [35]. The M5P tree is built recursively from the root node to the leaves, subdividing the space each time into smaller subsets. When pure subsets are reached the tree construction stops. Figure 12 in Section 4 includes an example of the regression trees built in this research.

Third, ensembles for regression [36] were also tested. Ensembles train a set of base regressors and combine their predictions in their final result. Bagging [37] is a kind of ensemble in which all the regressors belong to the same type (e.g. all are random trees) and they make different predictions because they are trained with different datasets (different samples of the original dataset) and the final prediction is computed as the average prediction of the base regressors. In this study, Random Forest [38] was used, a special case of Bagging, that uses random trees as base classifiers. Random trees are a variant of decision trees, with the difference that only a random subset of the attributes is considered when selecting the decision for a node.

Finally, another two regressors were considered as baseline methods for comparison with previous similar works [26,39]: the naïve approach and the regression model. The naïve approach, called ZeroR in Weka, is based on using the mean value of a parameter as a prediction of the parameter value (4.84 mm for the critical depth and 31.90 Hz for the chatter frequency in this case study). In other words: the naïve-approach prediction is always the same irrespective of the values of the process inputs. As a consequence, the error of any of the machine learning models should be significantly lower than the error of the baseline method, to assure that the machine learning model can predict new situations, in terms of machine position and feed direction. The second baseline that can be considered is a linear regression model, because this kind of simple model is usually able to capture the main relationship between inputs and outputs, at least in some input ranges. Linear Regression bases its prediction on a linear combination of the process inputs that minimizes the squared error of the training dataset.

4. Experimental validation of the proposed methodology

Firstly, this section presents the experimental validation of the proposed strategy, to which end the machine tool frequency response function (FRF) and the transfer function were both measured at the Tool Center Point (TCP). Then the analytical calculation of chatter frequency and the critical depth obtained from the FRFs with equations (4) and (5) are presented. Finally, the results of the machine-learning modeling of both variables introduced in section 3 are presented and the performance of each technique is discussed.

4.1 Machine description and FRFs measurement

The proposed strategy was applied to a Nicolas Correa Fox milling machine. The dynamic flexibility of the machine was measured at the Tool Center Point (TCP) in different positions that represent the transfer function in equation 2. This bridge milling machine has a frame that moves in the Y direction and a ram that slides vertically. The X movement is given by the work table that moves in the X direction. The machine can move 3000 mm in the Y direction and 1500 mm in the Z direction. The range of movement in the X direction is 5000 mm.

The instrumentation used to measure the Frequency Response Functions (FRFs) were: a dynamometric Dytran 5802A hammer with a nominal sensitivity of 0.25 mV/N, which incorporates a load cell for measuring the applied force; a PCB 356A32 triaxial piezoelectric accelerometer, with a nominal sensitivity of 100 mV/g, to measure the three Cartesian components of acceleration; and an OR35 model Analyzer with NVGATE software (from OROS). The FRFs were calculated by means of the H1 estimator method, with a bandwidth of 200 Hz. For each FRF measurement, three impacts were applied to the TCP and the average of these measurements was recorded as the result. The coherence of these three measurements was monitored to ensure a good coherence especially in natural frequencies. The FRF were measured with 3,200 lines, offering a resolution of 65 mHz.

The dynamic response at the TCP will change with the machine positions, because the natural machine frequencies and modal vectors are dependent on the machine position. In all, 27 positions were studied which represent all the working positions of the machine cube.

The milling machine bridge adds the Y and Z linear movement to the machine, and the working table adds the X linear movement, so only 18 measurements are needed to obtain the 27 dynamic responses: 9 measurements on the XZ plane and another 9 measurements on the XY plane (see figures 4 and 5). After these measurements are done, the sum of the different responses must be added up to obtain the 27 position responses. Figure 4 shows the 27 positions of the milling machine bridge:

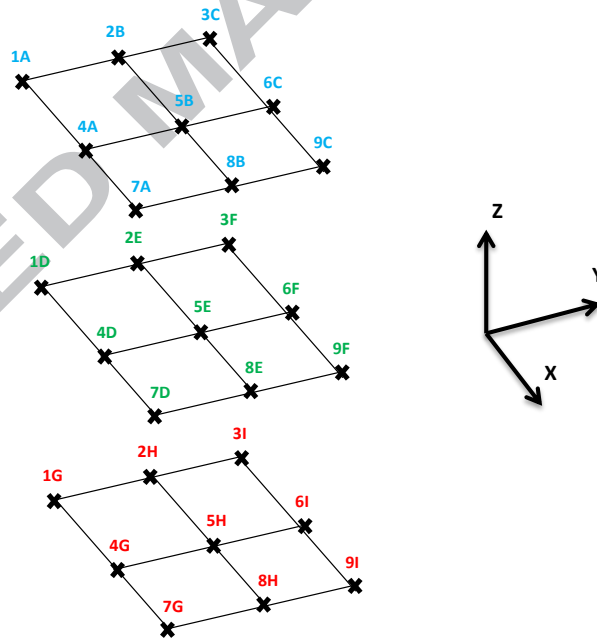


Figure 4. The 27 positions of the Milling bridge machine.

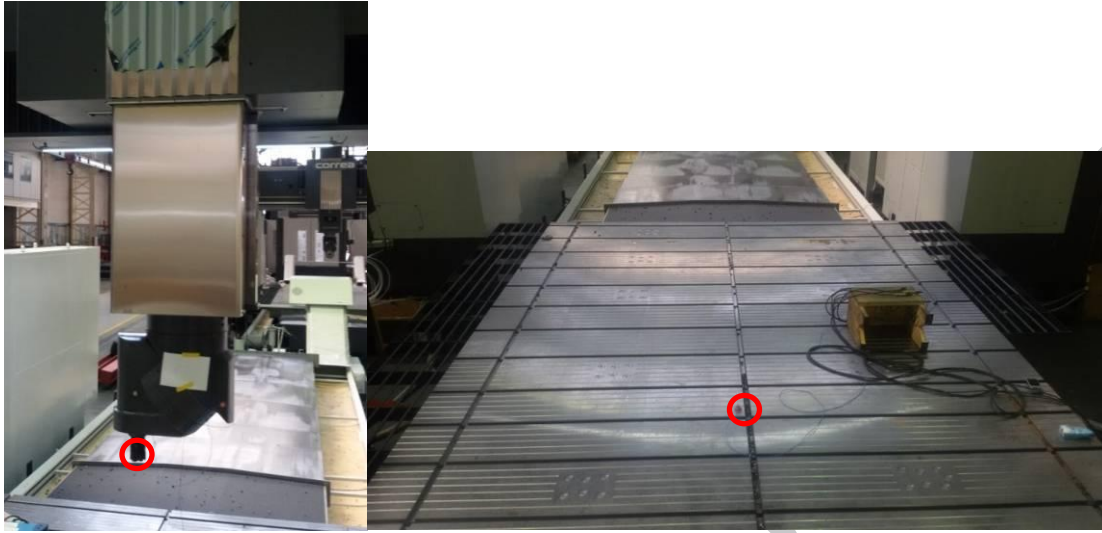


Figure 5. “E” position of the milling machine and position 5 of the working table

To obtain the 27 responses in the working positions the responses must be summed up:

- Bridge A, D and G position with working table positions 1, 4 and 7 → 9 positions: A1, A4, A7, D1, D4, D7, G1, G4 and G7
- Bridge positions B, E and H with working table positions 2, 5 and 8 → 9 positions: B2, B5, B8, E2, E5, E8, H2, H5 and H8
- Bridge positions C, F and I with working table positions 3, 6 and 9 → 9 positions: C3, C6, C9, F3, F6, F9, I3, I6 and I9.

An artificial excitation of the structure is needed at the TCP, in order to measure the dynamic response at the TCP experimentally, the frequency response at the TCP, and the input forces.

4.2 Analytical calculation of chatter frequency and critical depth

The modes that limit the productivity of the machine for a selected milling operation were identified. The representative milling operation was selected: the face milling of AISI 1045 steel in down-milling, with a radial immersion of 80%. Likewise, a MITSUBISHI SE445-125-B09 R tool was selected, with a diameter of 125 mm, fitted with nine carbide coated flutes ($Z=9$), reference SEER 1203 AFEN-JS F620, with a lead angle of 45° . The recommended rotational speed range for the tool is between 400 and 600 rpm, in accordance with similar works about face milling operations [42,43]. A tangential cutting coefficient of $K_t=1685 \text{ N/mm}^2$ was experimentally characterized for the cutting operation. The left sides of Figures 6, 7 and 8 in the polar charts show the critical cutting depths that separate chatter appearance from stability cutting conditions, which were obtained from equation (5). These critical depths were calculated for the whole range of machining directions within the $X_{MH}Y_{MH}$ working plane, thereby achieving the aforementioned polar charts. On the right sides of Figures 6, 7 and 8, the chatter frequencies obtained from equation (4) are shown, again in polar charts, taking once again the machine XY plane as the

diagram plane for any possible feed direction on that XY plane. The polar diagrams reflect each dynamic response of the bridge milling machine in a concordance roughing operation on F-114 steel with a 125 mm diameter mill fitted with nine cutters at 45°. Figures 7-9 show nine examples of the polar diagrams for the cutting depth and the chatter frequency in 9 representative machine working positions (A1, A2, A7, E2, E5, E8, I3, I6 and I9) included in Figure 4. The first 3 diagrams in Figure 6 show strong and sharp changes, both in depth of cut and chatter frequency depending on the feed direction. The following 6 diagrams in Figure 7 show a more stable behavior of the chatter frequency with few sharp changes and, in Figure 8, without any significant change, but with a clear but slight influence of the feed direction on the depth of cut.

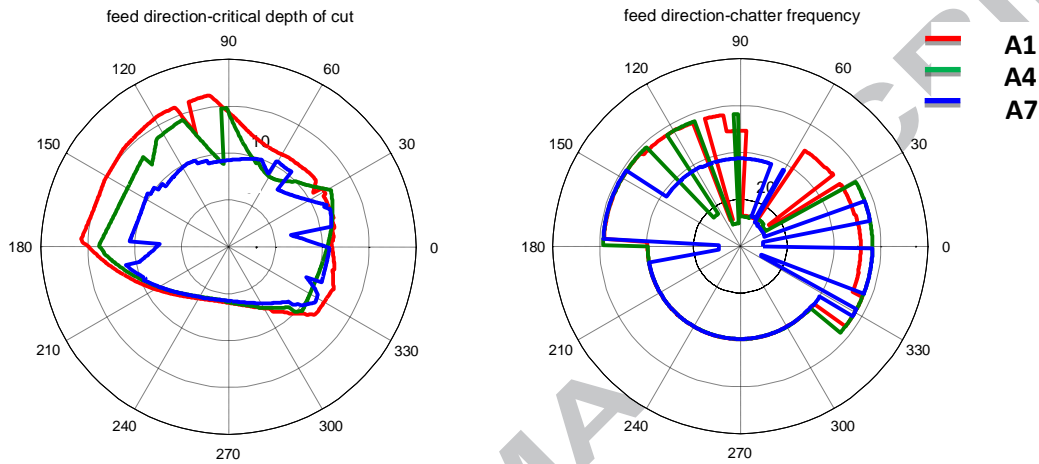


Figure 6. Polar diagrams for positions A1, A4, A7

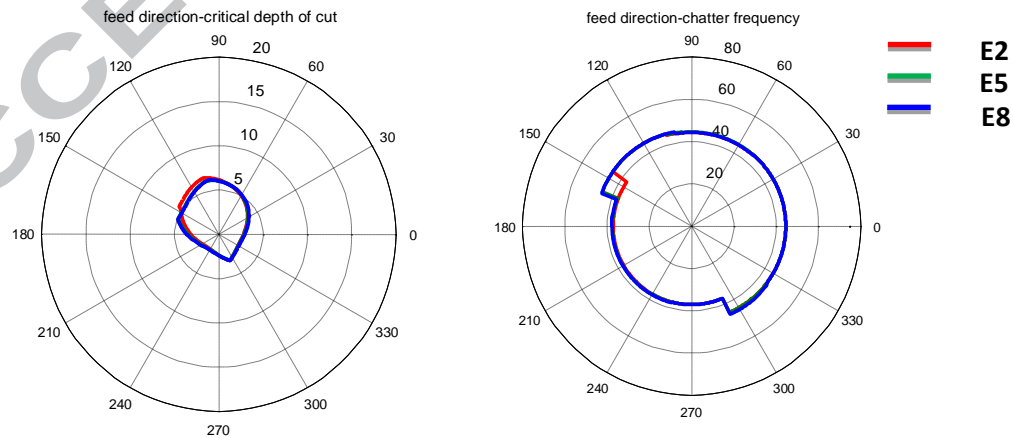
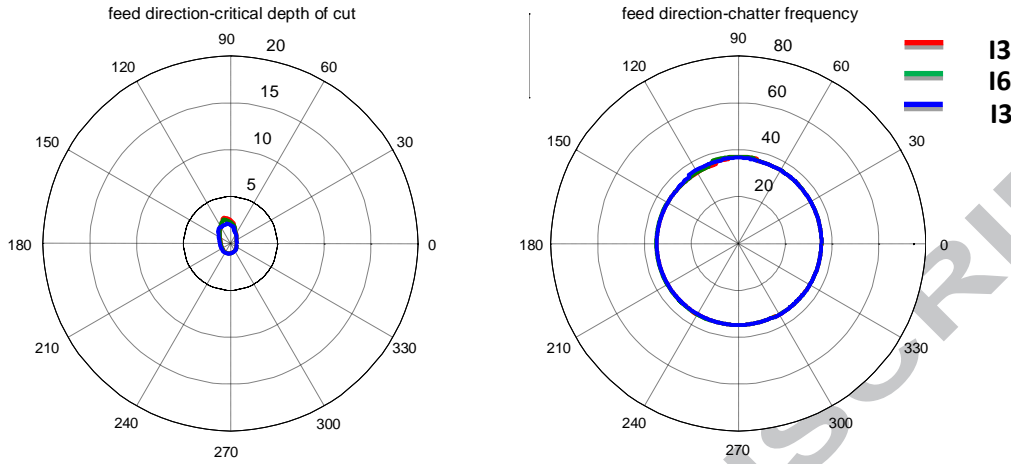


Figure 7. Polar diagrams for positions E2, E5, E8**Figure 8.** Polar diagrams for positions I3, I6, I9

In Figures 6-8, a significant difference can be seen between the polar diagrams of critical depth of cut and the polar diagrams of chatter frequency. The curves have progressive values in the critical depth-of-cut diagrams, while diagrams of the curves of the chatter frequencies can suddenly change value. These differences are because chatter frequencies are associated with the natural frequencies of the machine that are excited, which can vary abruptly between two adjacent machine positions, in contrast with the critical depths of cut values, where the curves are continuous values, which means that the milling machine will have progressive productivity limits while machining.

4.3 Machine-learning modelling

The dataset was generated from the polar diagrams calculated in Section 4.2 for the two outputs: the critical depth of cut (a_{pcrit}) and the chatter frequency (ω_c). Each dataset includes four inputs: three of them identify the machining position (X , Y and Z) and the forth is the feed direction (ϑ). All the inputs and outputs are continuous variables. The machine work space was divided into 27 different equi-spaced positions, previously presented in Section 4.1 and the feed direction was considered every 10° from 0° to 350° to cover any possible milling direction. Therefore, the datasets included 27 different milling directions with 36 feed directions for each machining position, forming a dataset of 972 instances. Table 1 summarizes the input and outputs, their units and the range of values presented in the dataset; the output variables are shown in bold.

Variable	Abbreviation	Range	Units
X position	X	0 - 5.000	mm
Y position	Y	0 - 2.370	mm
Z position	Z	0 - 1.230	mm
Feed direction	ϑ	0 – 350	$^\circ$
Chatter frequency	ω_c	1.14 - 65.25	Hz

Critical depth of cut	a_{pcrit}	0.68 - 19.20	mm
-----------------------	-------------	--------------	----

Table 1. Dataset variables and their variation range.

As presented in Section 3, the prediction of the chatter frequency and the critical depth are two regression problems from the artificial intelligence point of view and the machine-learning techniques are known as regressors. In both cases the dataset is split in two groups: the training set -that includes the instances that will be used to train the model- and the validation set -that includes the rest of the instances of the dataset that are used to measure the accuracy of the model-. The instances used at the validation stage were not used during the training-stage as is explained later. In this way, the prediction model will not be over-optimistic, as it should tend to be if the model is tested using the same instances as in the training stage. In this way the generalization capabilities of the prediction model in relation with new instances is considered.

The quality of the predictions is measured usually in manufacturing problems [21-23] with the Root Mean Square Error (RMSE). RMSE, as equation 6 shows, is calculated as the square root of the sum of squares of the prediction errors for each instance in the validation group divided by the number of instances in this group. Although RMSE is not a measure with a physical meaning from the point of view of the variable that it measures, it penalizes the models that provide a prediction that is very far off the real value in some validation instances, as it considers the squares of the errors and not the errors itself. Therefore, in the case of this research, RMSE is the most suitable measure, because a prediction far away from the real value will mean, for sure, an incorrect prediction in terms of machine performance (i.e. it will wrongly predict the chatter frequency that will lead to the appearance of chatter).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y)^2}{n}} \quad (6)$$

The following regressors, presented in Section 3, were tested using the WEKA data-mining tool [33]: regression trees, Random Forest and Multilayer Perceptrons. Besides, two baseline methods, ZeroR and Linear regression, were tested as baseline methods.

A 10-fold cross-validation repeated 10 times [44] was used to train and to validate the regressors, because it is the most popular validation schema. In 10-fold cross-validation, the original dataset is randomly divided into 10 datasets of equal-size called folds; nine of these datasets are used to train the prediction model and the last one is used to validate the model. In this way the model's accuracy is measured on a dataset different to the one that has been used to train the model. This process was repeated 9 times, to avoid the influence of the randomly selected instances for the validation dataset, using in each of the iterations, a different dataset for validation from the 10 folds in the first step of the process. Finally, this process was repeated ten times with different randomly split datasets, so the model accuracy is the average obtained from a hundred models. In this way, the variance of the model accuracy is reduced and the results of the machine learning model may be generalized, because the model is tested on new data that have not been used during the learning stage, thereby assuring its predictive capability in the face of unforeseen situations [44].

Some of the regressors in the tests had different parameters that should be fine-tuned to find their optimal values. A grid search was performed on the main parameters of each prediction model to find the right combination; the grid steps were fixed to 0.1 and only reduced to 0.05, if the accuracy of the model changes more than 10% in any of the programmed steps. In a grid test, all the considered values are tested against each other, as in a full-factorial experiment, as

opposed to a Taguchi experiment, for example, where only some combinations are tested [39]. Due to the small size of the dataset, the grid test can be performed on the 10-fold cross-validation, because the execution time is not significant. Table 2 collects the ranges and number of steps in which the parameters were varied. The default values in Weka [33] were taken for the parameters not listed in Table 2 and, therefore with negligible influence on the accuracy of the regressor. In the case of the MLPs, 210 different configurations were tested, while in the regression trees only the minimum number of instances per leaf were tuned, and in the Random Forest no-parameter was tuned. It is important to underline that MLPs have 3 parameters to tune while the regression tree has only one and Random Forest has none. This factor is very important in terms of tuning times and the need for an expert to perform the tuning process (definition of tuning limits, steps of the grid search, etc).

Prediction Model	Parameter	Range	Steps
MLPs	Momentum	0.1 to 0.5	7
MLPs	Learning rate	0.05 to 0.5	10
MLPs	Hidden layers	2,4,6	3
M5P Tree	Minimum number of instances	2,4,8,16	4

Table 2. Parameters of each regressor that are tuned and variation range and steps.

Table 3 shows the results for the best configuration of the tested machine learning techniques and baselines methods. MLPs are separated by the number of hidden layers from two to six. The number in brackets in Table 3 indicates the number of hidden layers in the MLPs and the minimum number of instances per leaf that provide the higher accuracy in the regression trees. The best results for MLPs are obtained for 6 hidden layers for critical depth a_{pcrit} (Momentum = 0.1 and Learning Rate = 0.05) and 4 hidden layers for the chatter frequency ω_c (Momentum = 0.5 and Learning Rate = 0.05). This table also compiles the computational time needed for a computer equipped with an Intel Core i5 2300 at 2.8 GHz processor to build each kind of model. Table 3 shows that all the machine learning models under study are more precise than the two considered baselines: ZeroR and linear regression, with significant statistical differences considering the corrected resampled t-test [45] with the significance level set at 5%. Regression trees are more precise than the best MLP configuration, but with no statistical difference, while Random Forest, the most accurate model, showed statistical differences with the other machine learning models; the difference with the best MLP was around 75% in the RMSE at the critical depth, a_{pcrit} , and around 30% for the chatter frequency, ω_c . In both cases, the RMSE for the Random Forest model are acceptable from the industrial point of view: around 0.95 mm for the critical depth and 7.3 Hz for the chatter frequency in the whole working and all the feed directions, taking into account that the critical depth varies from 0.68 mm to 19.20 mm and the chatter frequency from 1.14 Hz to 65.25 Hz in this studied range. Finally, if the computational time is considered, the quickest machine-learning models were the regression trees, while there were no significant differences between the time taken by Random Forest and the smallest MLPs (2 hidden layers) and the more accurate MLPs (4-6 hidden layers) required at least twice the computational time of Random Forest. If we consider that 210 different configurations of MLPs were built as opposed to only one configuration of Random Forest to find the most accurate model, the difference in computational time becomes very important. This advantage of ensembles compared with ANNs has been previously presented for other manufacturing tasks like roughness or wear prediction [24,26].

	Critical depth a_{pcrit}	Chatter frequency ω_c
--	----------------------------	------------------------------

	RMSE (mm)	Time (s)	RMSE (Hz)	Time (s)
ZeroR	3.8322	0	15.4575	0
Linear regression	2.0000	0	15.1136	0
MLPs [2]	1.796	0.25	10.344	0.35
MLPs [4]	1.777	0.43	9.388	0.43
MLPs [6]	1.689	0.64	9.426	0.62
M5P Tree [16]	1.3156	0.05	9.0982	0.02
Random Forest	0.9641	0.16	7.2484	0.32

Table 3. Precision of the tested regressors.

The MLP parameters momentum and learning rate both influenced model accuracy. Figure 9 shows the RMSE of the depth of cut models depending on their momentum and learning rate for the case of 6 hidden layers (the most accurate case). It can be concluded that the influence of the momentum is smaller than the influence of the learning rate, especially for low values of the learning rate, where the models are more accurate; the influence of the learning rate is almost linear in the range 0.15-0.5, but at values below 0.15 it becomes unstable. This behavior cannot be easily predicted and the assistance of an expert in ANN tuning is necessary to find the most accurate configuration.

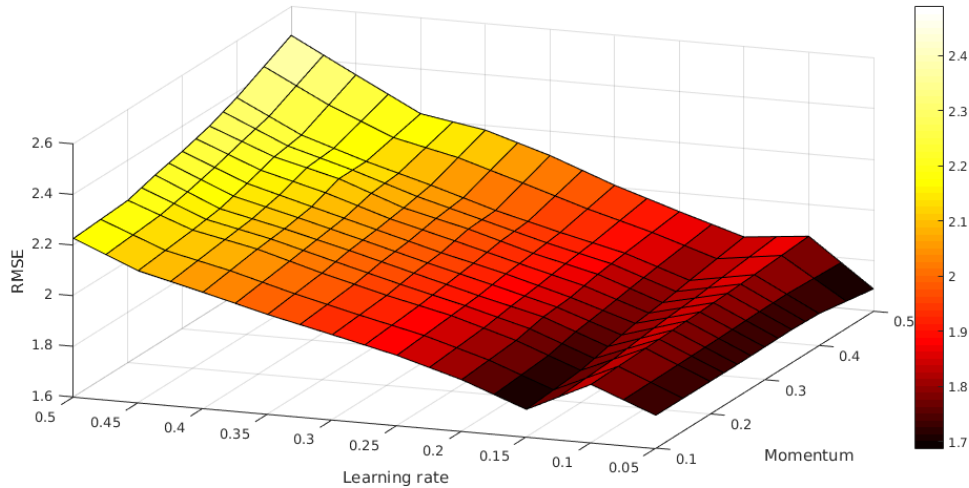


Figure 9. RMSE of the MLP models for the depth of cut for 6 hidden layers depending on the momentum and learning rate values.

Figure 10 shows the RMSE of the chatter frequency models depending on their momentum and learning rate for the case of 4 hidden layers (the most accurate case). In this case, the important influence of both parameters is able to halve the RMSE of the model from the worst to the best configuration; again, the influence of the momentum is smaller than the influence of the learning rate, especially for low values of the learning rate, where the models are more accurate; the influence of the learning rate is almost linear in the entire tuning range 0.05-0.5.

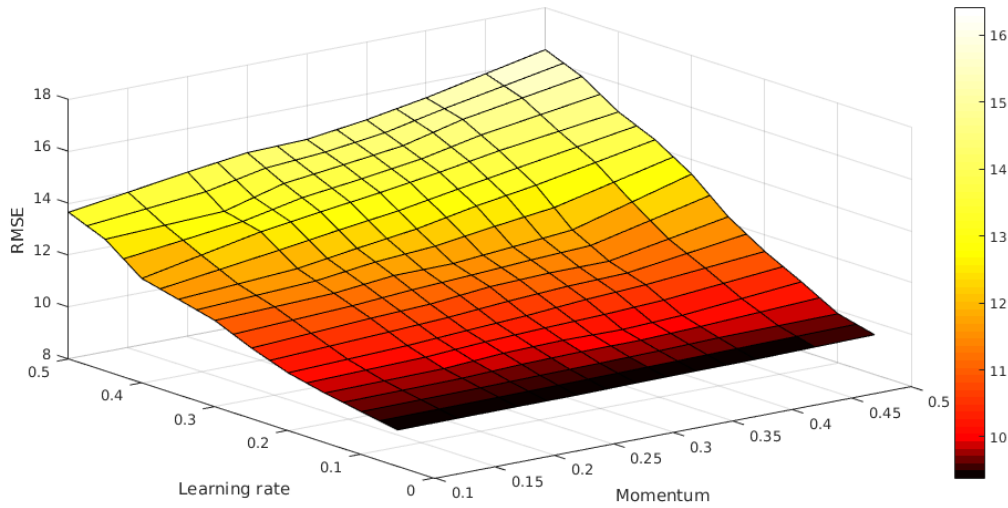


Figure 10. RMSE of the MLP models for the chatter frequency for 4 hidden layers depending on the values of both momentum and learning rate.

Finally, although regression trees do not achieve as high an accuracy as the Random Forest ensembles, as shown in Table 3, they are only 30% less accurate than Random Forest ensembles and they have the advantage of immediately providing visual information to the design engineer on machine critical performance, as other works have outlined for other manufacturing tasks such as wear prediction [26]. Figure 11 shows the regression tree built with all the dataset for the chatter frequency. An obvious conclusion is that there is a linear model LM1 ($\omega_c = -0.0001*Y + 0.0045*Z + 0.0144*\vartheta + 32.8437$) that predicts the chatter frequency directly in some area of the machine: if the Y-axis position is below 592 mm. As the Y-axis length of this machine is 3000 mm, the chatter frequency is directly defined by LM1 equation in 1/5 of the working volume, where both the Y and the Z axis positions have a low or negligible influence and only the ram position (Z-axis) and, obviously, the feed direction, ϑ , play a central role. If the machine designer wishes to increase the chatter frequency to improve the machine performance, then the dynamic stiffness of the ram should be increased. In another 2/5 of the working volume, that is: if the Y-axis position is above the limit of 592 mm and the Z-axis position is above 1.777 mm, then all that needs to be done is to generate another two linear models, LM8 and LM9, to predict the chatter frequency. The feed direction only plays a role in the medium region of the Y-Axis, and the influence of the X-axis position is very limited, because only two branches of the tree requires any value of this parameter. This information can be used to analyze why the Y-axis is so critical in chatter frequency for the milling machine under study and which, at first glance, are the worst conditions.

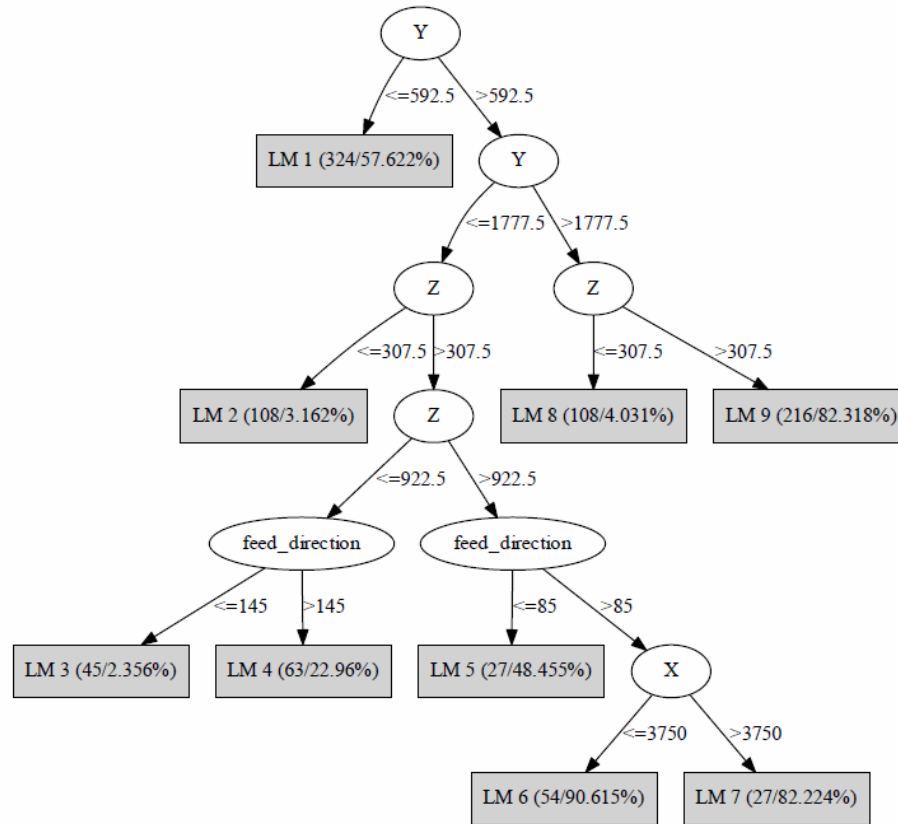


Figure 11. Regression tree for the prediction of chatter frequency

5. Conclusions

The productivity of milling operations is mainly restricted by the appearance of self-induced vibrations, especially regenerative chatter vibrations. Although there is an extensive bibliography on chatter suppression for regenerative vibrations of tools and workpieces, there are few studies on chatter related to the machine component modes of low dynamic stiffness. Two key parameters are linked to these vibrations: the depth of cut that may be achieved without those regenerative vibrations and the frequency of the vibrations, also known as the chatter frequency. In this research, the calculation of both parameters is done by means of a 3-step strategy that combines experimental FRFs recorded by the Tool Center Point (TCP) that produce data for analytical calculations of both parameters, and various machine-learning techniques by which the calculations may be used to build prediction models for any combination of milling direction and machine position. This strategy has not previously been presented in the bibliography, where mechanical simulations are rarely combined with machine-learning techniques to provide high accuracy models that can be industrially validated.

Accordingly, the mechanical performance of the milling machine was firstly characterized at discrete positions on the working volume by means of experimental frequency responses at the TCP. Combined with mechanical characterizations of the tools and analytical models of milling forces and dynamic chip thicknesses, the discrete calculation of critical depths of cut and chatter frequencies may be done throughout the working volume. The milling tests involved concordance roughing on F-114 steel with a 125 mm diameter mill fitted with nine cutters aligned at 45°.

Then, various machine-learning techniques have been tested to predict these two parameters. The following regression techniques have been tested: artificial neural networks, regression trees and random forest. The results have shown that regression trees are more precise than the best MLP configuration, but that there is no statistical difference between either, while Random Forest is the most accurate model with a statistical advantage over the other machine learning models; the difference with the best MLP was around 75% in the RMSE for the critical depth a_{pcrit} and 30% for the frequency ω_c . In both cases, the RMSE for the Random Forest model were acceptable from the industrial point of view: around 0.95 mm for the critical depth and 7.3 Hz for the chatter frequency in the whole working volume and in all the feed directions, taking into account that the critical depth varied from 0.68 mm to 19.20 mm and the chatter frequency from 1.14 Hz to 65.25 Hz in the range under study. Besides, if the computational time is considered, the quickest machine-learning models were the regression trees and the Random Forest ensembles, while the most accurate MLPs were significantly slower, especially considering that the tuning of the three parameters of the MLP requires the testing of hundreds of possible configurations of MLPs.

Finally, it may be said that regression trees can be incorporated in interesting and highly accurate tools that quickly provide useful information to the design engineer on the critical performance of machines. In this study case, the Y-axis was the only axis of the milling machine that strongly influenced chatter frequency, while the influence of the Z-axis position was lower, and the X-axis had a very limited influence, because only two branches of the tree required any value for this parameter. This information can be used to analyze why the Y-axis is so critical in chatter frequency (e.g. by strengthening the design of the saddle and the mechanical coupling between the bridge and the ram) and which, at first glance, are the worst conditions.

Further research will focus on the analysis and the application of machine-learning techniques, especially to unbalanced datasets with the sort of industrial problem discussed in this work, due to the extreme and unique situations where chatter tends to generate very unbalanced datasets. Moreover, online measurement strategies will be applied for the characterization of the dynamic performance of the milling machine under actual working conditions. This strategy will solve the present limitation that requires the use of cutting tests and modal analysis to arrive at ad-hoc characterizations of dynamic machine performance.

Acknowledgments

This investigation was partially supported by the Hidrodamp Project (IDI-20110453) of the Spanish Ministry of Economy and Competitiveness. The research work was facilitated by support from Nicolás Correa S.A., which provided the heavy-duty milling machine used in the experimental tests. The authors would especially like to thank Mr. Javier Hernando, from Nicolas Correa Company and Alvar Arnaiz from the University of Burgos, for their kind-spirited and useful advice.

References

- [1] P. Nieslony, G.M. Krolczyk, S. Wojciechowski, R. Chudy, K. Zak, R.W. Maruda, Surface quality and topographic inspection of variable compliance part after precise turning, *Applied Surface Science*, 434, 2018, 91-101.
- [2] S. Wojciechowski, K. Mrozek, Mechanical and technological aspects of micro ball end milling with various tool inclinations, *International Journal of Mechanical Sciences*, 134, 2017, 424-435.

- [3] S. Wojciechowski, R. W. Maruda, G. M. Krolczyk, P. Niesłony, Application of signal to noise ratio and grey relational analysis to minimize forces and vibrations during precise ball end milling, *Precision Engineering*, 51, 2018, 582-596.
- [4] J. Tlustý, M. Poláček, The stability of machine tool against self-excited vibrations in machining, *International Research in Production Engineering*, 1963, 465–474.
- [5] G. M. Krolczyk, J. B. Krolczyk, R. W. Maruda, S. Legutko, M. Tomaszewski, Metrological changes in surface morphology of high-strength steels in manufacturing processes, *Measurement: Journal of the International Measurement Confederation*, 88, 2016, 176-185.
- [6] Y. Altintas, C. Brecher, M. Weck, S. Witt, Virtual machine tools, *Annals of the CIRP*, 54, 2005, 651-674.
- [7] G. Catania, N. Mancinelli, Theoretical-experimental modeling of milling machines for the prediction of chatter vibration, *International Journal of Machine Tools & Manufacture*, 51(4), 2011, 339-348.
- [8] M. Uekita, Y. Takaya, Tool condition monitoring technique for deep-hole drilling of large components based on chatter identification in time–frequency domain, *Measurement: Journal of the International Measurement Confederation*, 103, 2017, 199-207.
- [9] G. Bolar, A. Das, S.N. Joshi, Measurement and analysis of cutting force and product surface quality during end-milling of thin-wall components, *Measurement: Journal of the International Measurement Confederation*, 121, 2018, 190-204.
- [10] S. S. Park, R. Rahnama, Robust chatter stability in micro-milling operations, *CIRP Annals*, 59(1), 2010, 391-394.
- [11] F. A. Khasawneh, E. Munch, Chatter detection in turning using persistent homology, *Mechanical Systems and Signal Processing*, 70-71, 2016, 527-541.
- [12] S. Wojciechowski, R.W. Maruda, S. Barrans, P. Nieslony, G.M. Krolczyk, Optimisation of machining parameters during ball end milling of hardened steel with various surface inclinations, *Measurement: Journal of the International Measurement Confederation*, 111, 2017, 18-28.
- [13] M. Reuss, A. Dadalau, A. Verl, Friction variances of linear machine tool axes, *Procedia CIRP*, 4, 2012, 115–119.
- [14] M. Lamraoui, M. El Badaoui, F. Guillet, Chatter detection in CNC milling processes based on wiener-SVM approach and using only motor current signals, *Mechanisms and Machine Science*, 23, 2015, 567-578.
- [15] E. Kuljanic, G. Totis, M. Sortino, Development of an intelligent multisensor chatter detection system in milling, *Mechanical Systems and Signal Processing*, 23 (5), 2009, 1704-1718.
- [16] I. Bediaga, J. Muñoa, J. Hernández, L.N. López de Lacalle, An automatic spindle speed selection strategy to obtain stability in high-speed milling, *International Journal of Machine Tools and Manufacture*, 49(5), 2009, 384–394.
- [17] Z. Yao, D. Mei, Z. Chen, On-line chatter detection and identification based on wavelet and support vector machine, *Journal of Materials Processing Technology*, 210(5), 2010, 713-719.
- [18] M. A. Sofuoglu, S. Orak, Prediction of stable cutting depths in turning operation using soft computing methods, *Applied Soft Computing Journal*, 38, 2016, 907-921.

- [19] J. Friedrich, C. Hinze, A. Renner, A. Verl, A. Lechler, Estimation of stability lobe diagrams in milling with continuous learning algorithms, *Robotics and Computer-Integrated Manufacturing*, 43, 2017, 124-134.
- [20] S. Hu, X. Huang, Y. Zhang, C. Lv, Reliability Analysis of the Chatter Stability during Milling Using a Neural Network, *International Journal of Aerospace Engineering*, 2016, art. no. 5259821.
- [21] J. Abellan-Nebot, F. Romero-Subirón, A review of machining monitoring systems based on artificial intelligence process models, *The International Journal of Advanced Manufacturing Technology*, 47, 2010, 237-257.
- [22] P. G. Benardos and G. C. Vosniakos, Predicting surface roughness in machining: a review, *International Journal of Machine Tools and Manufacture* 43(8), 2003, 833-844.
- [23] A. Siddhpura, R. A. Paurobally, Review of flank wear prediction methods for tool condition monitoring in a turning process, *The International Journal of Advanced Manufacturing Technology*, 65(1), 2013, 371-395.
- [24] D. Y. Pimenov, A. Bustillo, T. Mikolajczyk, Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth, *Journal of Intelligent Manufacturing*, 29(5), 2018, 1045-1061.
- [25] A. Bustillo, J.J. Rodriguez, Online breakage detection of multitooth tools using classifier ensembles for imbalanced data, *International Journal of Systems Science*, 45(12), 2014, 2590-602.
- [26] A. Bustillo, D. Y. Pimenov, M. Matuszewski, T. Mikolajczyk, Using artificial intelligence models for the prediction of surface wear based on surface isotropy levels, *Robotics and Computer-Integrated Manufacturing*, 53, 2018, 215-227.
- [27] L. Kuncheva, "Combining Pattern Classifiers: Methods and Algorithms: Second Edition", 2014, Wiley Inc. ISBN: 978-1-118-31523-1.
- [28] J. Yu, L. Xi, X. Zhou, Identifying source(s) of out-of-control signals in multivariate manufacturing processes using selective neural network ensemble, *Engineering Applications of Artificial Intelligence*, 22, 2009, 141-52.
- [29] S. Cho, S. Binsaeid, S. Asfour, Design of multisensor fusion-based tool condition monitoring system in end milling, *The International Journal of Advanced Manufacturing Technology*, 46, 2010, 681-94.
- [30] S. Binsaeid, S. Asfour, S. Cho, A. Onar, Machine ensemble approach for simultaneous detection of transient and gradual abnormalities in end milling using multisensor fusion, *Journal of Materials Processing Technology*, 209 (10), 2009, 4728-38.
- [31] T. Liao, F. Tang, J. Qu, P. Blau, Grinding wheel condition monitoring with boosted minimum distance classifiers, *Mechanical Systems and Signal Processing*, 22, 2008, 217-32.
- [32] J. J. Zulaika, F. Campa, L.N. Lopez Lacalle, An integrated process-machine approach for designing productive and lightweight milling machines, *International Journal of Machine Tools and Manufacture*, 51, 2011, 591-604.
- [33] M. Hall, E. Frank, G. Holmes G, B. Pfahringer, P. Reutemann, I.H. Witten, *The WEKA Data Mining Software: An Update; SIGKDD Explorations*, 11(1), 2009, 10-18.
- [34] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, Inc. New York, NY, USA, 1995.

- [35] J. R. Quinlan, Learning with continuous classes, Proc. of the 5th Australian Joint Conference on Artificial Intelligence, vol. 92, Singapore, 1992. 343–348.
- [36] J. Mendes-Moreira, C. Soares, A.M. Jorge, J.F. De Sousa, Ensemble approaches for regression: A survey, ACM Computing Surveys, 45(1), 2012, art. no. 10.
- [37] L. Breiman, Bagging predictors, Machine learning, 24(2), 1996, 123–140.
- [38] L. Breiman, Random forests, Machine Learning, 45(1), 2001, 5–32.
- [39] D. Teixidor, M. Grzenda, A. Bustillo, J. Ciurana, Modeling pulsed laser micromachining of micro geometries using machine-learning Techniques, Journal of Intelligent Manufacturing, 26(4), 2015, 801-814.
- [40] F. J. Campa, L. N. López de Lacalle, A. Lamikiz, and J. A. Sánchez, Selection of Cutting Conditions for a Stable Milling of Flexible Parts with Bull-Nose End Mills, Journal of Materials Processing Technology 191 (1–3), 2007, 279–282.
- [41] C. M. Giorgio Bort, M. Leonesio, P. Bosetti, A model-based adaptive controller for chatter mitigation and productivity enhancement in CNC milling machines, Robotics and Computer-Integrated Manufacturing, 40, 2016, 34-43.
- [42] D.Y. Pimenov, V.I. Guzeev, T. Mikolajczyk, K. Patra, A study of the influence of processing parameters and tool wear on elastic displacements of the technological system under face milling. International Journal of Advanced Manufacturing Technology, 92(9-12), 2017. 4473-4486.
- [43] D.Y. Pimenov, V.I. Guzeev, A.A. Koshin, Elastic displacement of a technological system in face milling with tool wear. Russian Engineering Research 31(11), 2011, 1105-1109.
- [44] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (San Mateo, CA: Morgan Kaufmann), 2 (12), 1995, 1137–1143.
- [45] C. Nadeau, Y. Bengio, Inference for the generalization error, Machine Learning, 52(3), 2003, 239–281.

Highlights:

- Presentation of an extensive dataset of structural chatter frequency and amplitude
- Presentation of a new approach to predict structural chatter
- Experimental validation of the new approach on the extensive dataset
- Random forest ensembles provide the highest accuracy and are more easily optimized
- Regression trees provide immediate visual information to the design engineer

